#1

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Michael Wall
Serial No.:	10/043,377
Filed:	January 11, 2002
For:	Recursive Categorical Sequence Assembly
Group Art Unit:	2673
Examiner:	Unknown
Attorney's Docket No.	N8837
Customer No.	23456

PRELIMINARY AMENDMENT

Box Missing Parts
Commissioner for Patents
Washington, DC 20231

Dear Sir:

Applicant is filing this Preliminary Amendment with respect to the above captioned patent application.

CLEAN VERSION OF REPLACEMENT PARAGRAPHS AND CLAIMS

Please amend the specification as follows. A marked-up version of the amended specification appears in the section entitled "**VERSION WITH MARKINGS TO SHOW CHANGES MADE**" following this amendment.

In the Specification:

Please amend the specification as follows:

Please delete the paragraph beginning on page 10, line 14 with "In each read..." and ending on page 10, line 21 with "...as similar methods." and replace with the following:

In each read there may be some longest repeat. The repeat may be of a single base or a repetition of multiple bases. The algorithm may cluster each of the reads based on the longest repeat of a single base or the longest repeat of multiple bases. A repeating region is well known in the art. As an example, as shown in SEQ ID NO: 1, the sequence "TAGAGAGAGAGAGATCATCGAT" contains a GA repeat from bases three through sixteen, which is this sequence's longest repeat. The categories of the present invention may be based on the GC percentage or any other percentage of the repeat as well as similar methods.

Please delete the paragraph beginning on page 10, line 23 with "Each read may have..." and ending on page 11, line 3 with "...is a G or a C." and replace with the following:

Each read may have a long high or low entropy area within the read's sequence. This area will have some GC content. An example of a sequence with high entropy and high GC content is SEQ ID NO: 2, "GAGTGTATCTGCCCGCCGGCGTGCCCGGCTAC". The entropy is high because there is not an even distribution of A, G, C, and T. The GC percentage is high because over 70% of the sequence is a G or a C.

Please amend the application by adding to the application, as a separate part of the disclosure, the content of the paper copy of the sequence listing, filed herewith.

REMARKS

Pursuant to 37 C.F.R. 1.821(b), 37 C.F.R. 1.821(c), 37 C.F.R. 1.821(d) and M.P.E.P. 2422.02, Applicant submits this amendment to the specification in order to provide the sequence identifiers for the presented sequences. No new matter has been added to the application. If the examiner has any questions or concerns, it would be appreciated if he or she would telephone the undersigned.

The Commissioner is authorized to charge any deficiency or credit any overpayment in connection with this preliminary amendment to Deposit Account No. 23-0035.

Respectfully submitted,

Doug Schelling

Douglas W. Schelling, Ph.D.
Registration No. 48,335
WADDEY & PATTERSON
A Professional Corporation
Customer No. 23456
ATTORNEY FOR APPLICANT

VERSION WITH MARKINGS TO SHOW CHANGES MADE

In the Specification:

Please amend the specification as follows:

Please delete the paragraph beginning on page 10, line 14 with "In each read..." and ending on page 10, line 21 with "...as similar methods." and replace with the following:

In each read there may be some longest repeat. The repeat may be of a single base or a repetition of multiple bases. The algorithm may cluster each of the reads based on the longest repeat of a single base or the longest repeat of multiple bases. A repeating region is well known in the art. As an example, as shown in SEQ ID NO: 1, the sequence "TAGAGAGAGAGAGATCATCGAT" contains a GA repeat from bases three through sixteen, which is this sequence's longest repeat. The categories of the present invention may be based on the GC percentage or any other percentage of the repeat as well as similar methods.

Please delete the paragraph beginning on page 10, line 23 with "Each read may have..." and ending on page 11, line 3 with "...is a G or a C." and replace with the following:

Each read may have a long high or low entropy area within the read's sequence. This area will have some GC content. An example of a sequence with high entropy and high GC content is SEQ ID NO: 2, "GAGTGTATCTGCCCGCCGGCGTGCCCGGCTAC". The entropy is high because there is not an even distribution of A, G, C, and T. The GC percentage is high because over 70% [percent] of the sequence is a G or a C.

CERTIFICATE OF FIRST CLASS MAILING

I hereby certify that this Preliminary Amendment, paper copy of the sequence listing (1 page), and self addressed post card are being deposited with the United States Postal Service as first class mail in an envelope addressed to:

Box Missing Parts
Commissioner for Patents
Washington, DC 20231,

On October 1, 2002.

Douglas W. Schelling, Ph.D.

Doug Schelling

Registration No. 48,335

10-1-02

Date

Seq_listing.ST25

SEQUENCE LISTING

<110> Large Scale Biology Corporation

<120> RECURSIVE CATEGORICAL SEQUENCE ASSEMBLY

<130> 00801-0211-NPUS00

<140> 10/043,377
<141> 2002-01-11

<160> 2

<170> PatentIn version 3.1

<210> 1
<211> 22
<212> DNA
<213> Unknown

<220>
<223> Ficticious example provided on page 10 of specification.

<400> 1
tagagagaga gagatcatcg at 22

<210> 2
<211> 32
<212> DNA
<213> Unknown

<220>
<223> Ficticious example provided on page 11 of specification.

<400> 2
gagtgtatct gcccgccggc gtgcccggct ac 32